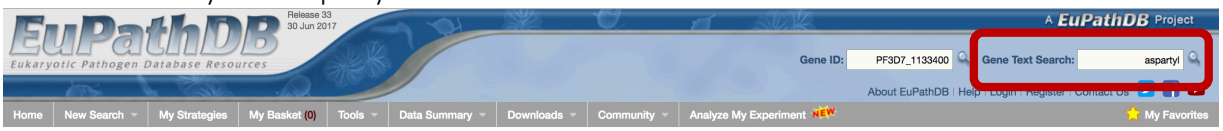
Finding Genes, Building Search Strategies and Visiting a Gene Page

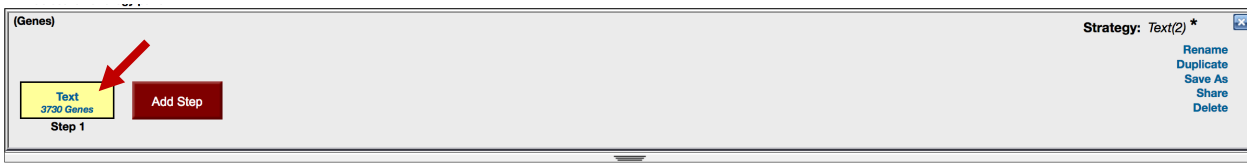
1. Finding genes that encode aspartyl proteases using text search.
   For this exercise navigate to http://eupathdb.org

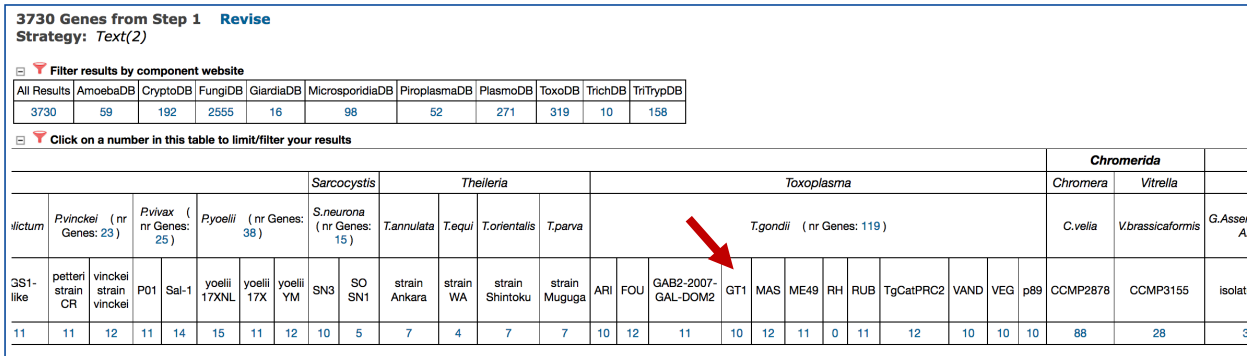
   a. Find all genes that contain the word aspartyl in the gene record.

   Hint: use the keyword aspartyl "Gene Text Search" box.



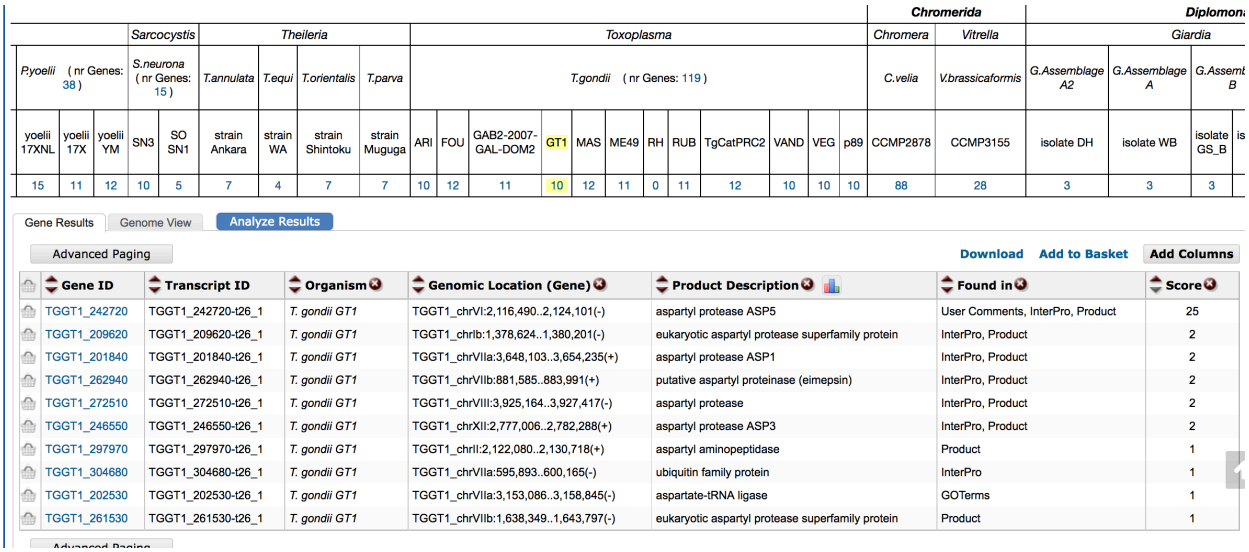
   - How many genes did you get?
   - Look closely at the different sections of the result page. How many of those are in *Toxoplasma gondii* strain GT1? How did you find this out?
     (*Hint* – the filter table is located between the strategy panel and the result table and shows the distribution of results across the organisms that you searched. Click on a number to only display results from a specific species or strain – you will have to scroll to the right of the page since the table is wide).





   b. Filter your results to only look at the results from *T. gondii* GT1.

   To filter your results, click on the number in the box under GT1



			Sarcocystis		Theileria				Toxoplasma											Chromerida		Diplomon		
																				Chromera	Vitrella	Giardia		
P.yoelii	(nr Genes: 38)		S.neurona (nr Genes: 15)		T.annulata	T.equi	T.orientalis	T.parva	T.gondii (nr Genes: 119)											C.velia	V.brassicaformis	G.Assemblage A2	G.Assemblage A	G.Assemt B
yoelii 17XNL	yoelii 17X	yoelii YM	SN3	SO SN1	strain Ankara	strain WA	strain Shintoku	strain Muguga	ARI	FOU	GAB2-2007-GAL-DOM2	GT1	MAS	ME49	RH	RUB	TgCatPRC2	VAND	VEG	p89	CCMP2878	CCMP3155	isolate DH	isolate WB	isolate GS_B	is
15	11	12	10	5	7	4	7	7	10	12	11	10	12	11	0	11	12	10	10	10	88	28	3	3	3	

Gene Results | Genome View | Analyze Results

Advanced Paging | Download | Add to Basket | Add Columns

Gene ID	Transcript ID	Organism	Genomic Location (Gene)	Product Description	Found in	Score
TGGT1_242720	TGGT1_242720-t26_1	T. gondii GT1	TGGT1_chrVI:2,116,490..2,124,101(-)	aspartyl protease ASP5	User Comments, InterPro, Product	25
TGGT1_209620	TGGT1_209620-t26_1	T. gondii GT1	TGGT1_chrIb:1,378,624..1,380,201(-)	eukaryotic aspartyl protease superfamily protein	InterPro, Product	2
TGGT1_201840	TGGT1_201840-t26_1	T. gondii GT1	TGGT1_chrVIIa:3,648,103..3,654,235(+)	aspartyl protease ASP1	InterPro, Product	2
TGGT1_262940	TGGT1_262940-t26_1	T. gondii GT1	TGGT1_chrVIIb:881,585..883,991(+)	putative aspartyl proteinase (eimepsin)	InterPro, Product	2
TGGT1_272510	TGGT1_272510-t26_1	T. gondii GT1	TGGT1_chrVIII:3,925,164..3,927,417(-)	aspartyl protease	InterPro, Product	2
TGGT1_246550	TGGT1_246550-t26_1	T. gondii GT1	TGGT1_chrXII:2,777,006..2,782,288(+)	aspartyl protease ASP3	InterPro, Product	2
TGGT1_297970	TGGT1_297970-t26_1	T. gondii GT1	TGGT1_chrII:2,122,080..2,130,718(+)	aspartyl aminopeptidase	Product	1
TGGT1_304680	TGGT1_304680-t26_1	T. gondii GT1	TGGT1_chrVIIa:595,893..600,165(-)	ubiquitin family protein	InterPro	1
TGGT1_202530	TGGT1_202530-t26_1	T. gondii GT1	TGGT1_chrVIIa:3,153,086..3,158,845(-)	aspartate-tRNA ligase	GOTerms	1
TGGT1_261530	TGGT1_261530-t26_1	T. gondii GT1	TGGT1_chrVIIb:1,638,349..1,643,797(-)	eukaryotic aspartyl protease superfamily protein	Product	1

Advanced Paging

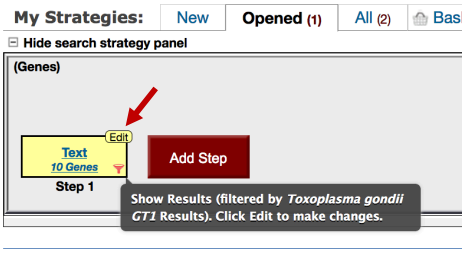
Notice how the number in the box become highlighted in yellow.

c. Examine your results. Specifically look at the Product Description column.

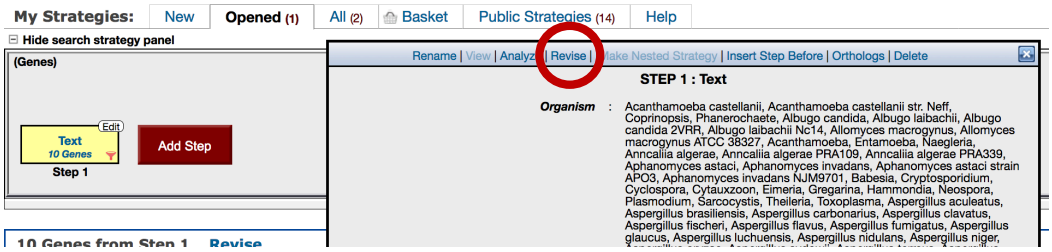
How many genes have the word aspartyl in their description?

d. Revising a search.

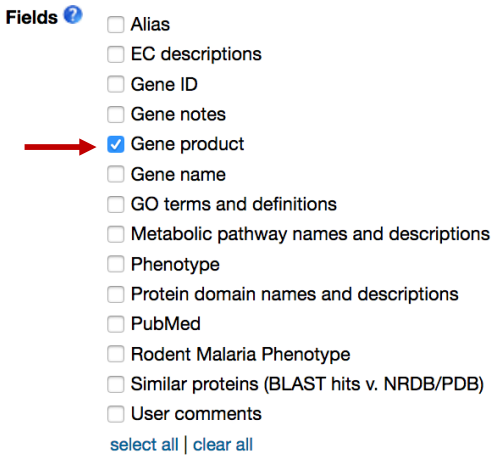
Revising a search allows you to change the parameters of a search. To revise a search move your cursor over the box in your search strategy until you see the edit link.



Click on the edit link then click on the "Revise" link

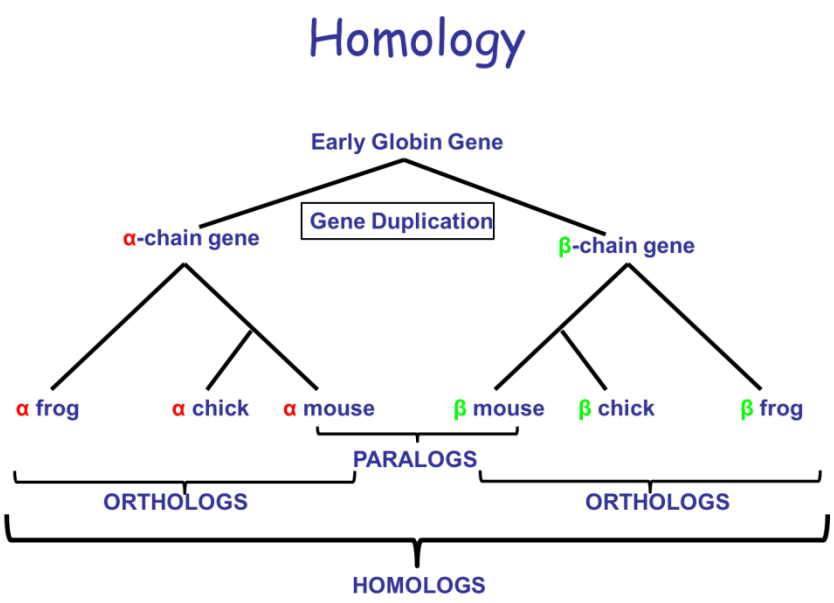


Once you click on revise, on the next popup page, scroll to the "Fields" section and only choose the field called "Gene Product"



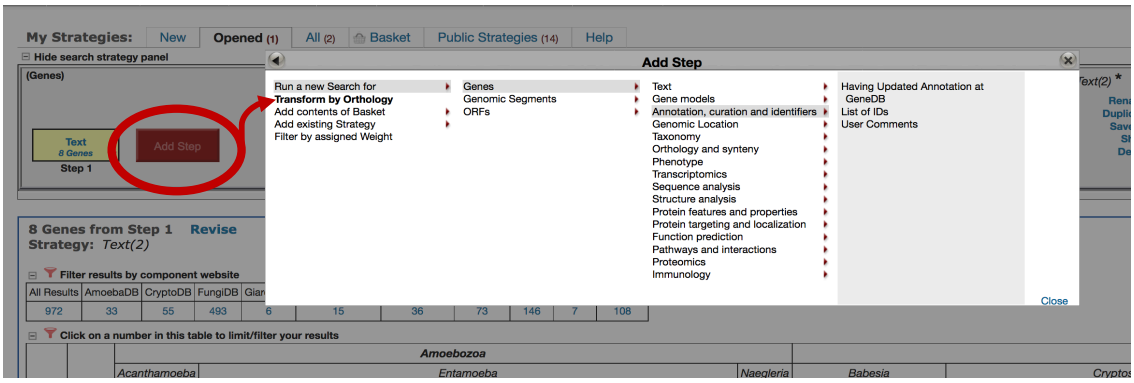
Next click on the "Run Search" button. How did your results change? Look at the product description column – do all genes have the word aspartyl in their descriptions?

2. Leveraging orthology to identify more aspartyl proteases (we will have a discussion on orthology)

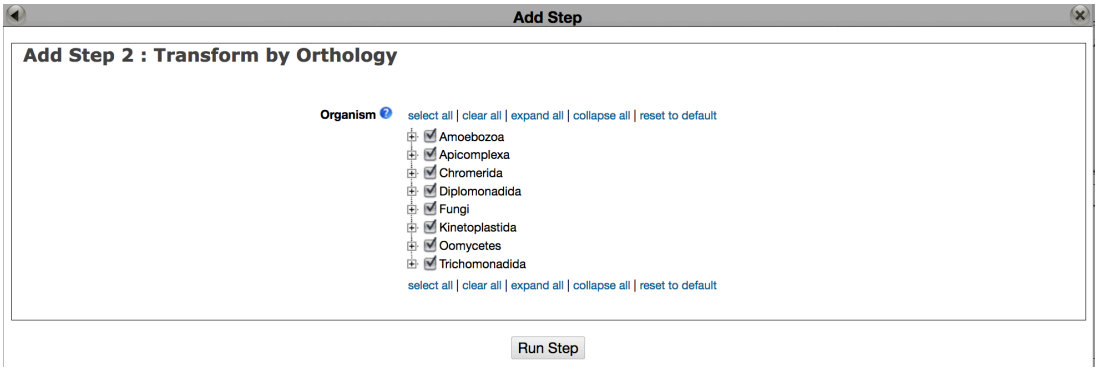


EuPathDB resources have a tool that allows you to transform a list of genes to their orthologs and paralogs across organisms in the database.

a. Starting with your results in part 1, click on the add step button (red button in the strategy panel).



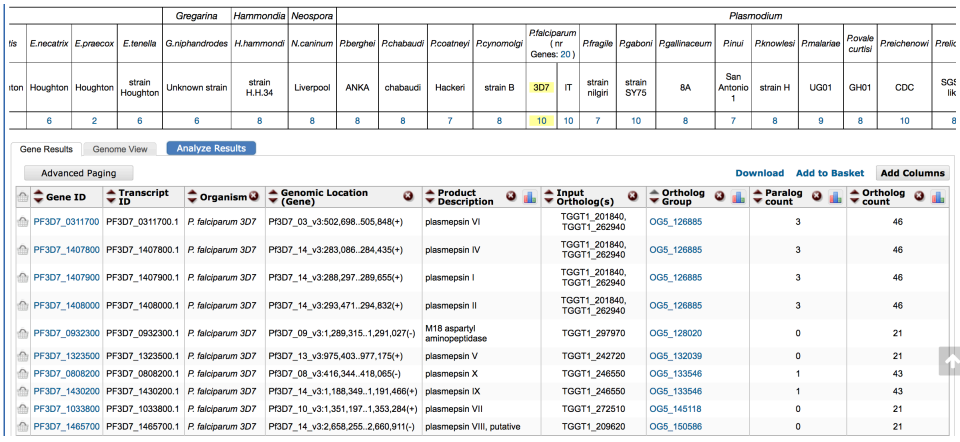
b. In the popup, click on the "Transform by Orthology" option.



Keep the default parameters selected and click on the Run Step button. Think about what are you asking the database to return.

c. What do you results look like? How many genes in total did you get? What is the distribution of orthologs across apicomplexan (hint: explore the filter table)?

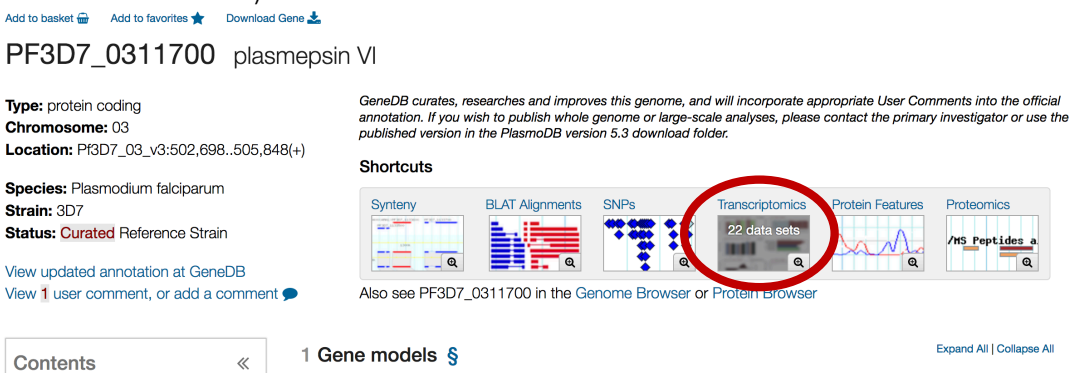
d. Filter your results to view genes from *Plasmodium falciparum* 3D7. What are the genes called?



3. Exploring specific aspartyl protease gene.

The column called “Gene ID” contains unique gene identifiers for each of the genes in your list. These IDs are also linked to the gene page.

a. Click on the gene ID for plasmepsin IV (ID: PF3D7_0311700).
b. Notice you are now on the gene page for plasmepsin IV in PlasmoDB
c. Try and determine at which stage of the plasmodium life cycle is this gene expressed. To do this go to the transcriptomics section of the gene page (hint: you can click on the transcriptomics short cut at the top of the page or you can scroll down the page until you find this section).



d. For example look at the first experiment in the list “Erythrocytic expression time series (3D7, DD2, HB3)”. Is this gene expressed in this time series?
e. Now, examine the experiment called “Strand specific transcriptomes of 4 life cycle stages”. Is this gene expressed at the stages represented in this experiment? Which stages?
f. Repeat this for each of the aspartyl proteases in *Plasmodium*. Are these proteases expressed at different stages?